



STUDENTS
AT
THE **CENTER**
A JOBS FOR THE FUTURE INITIATIVE

**MEASURING DEEPER LEARNING:
NEW DIRECTIONS IN FORMATIVE ASSESSMENT**

By Robert Rothman
July 2018

JOBS FOR THE FUTURE

JFF is a national nonprofit that drives transformation in the American workforce and education systems. For 35 years, JFF has led the way in designing innovative and scalable solutions that create access to economic advancement for all. Join us as we build a future that works.

WWW.JFF.ORG

STUDENTS AT THE CENTER

A JOBS FOR THE FUTURE INITIATIVE

Students at the Center is a leading national voice for the knowledge, skills, and dispositions all youth need to succeed in college, career, and civic life. Our work synthesizes—and adapts for practice—current research on key components of student-centered approaches to learning that lead to deeper learning outcomes. Our goal is to strengthen the ability of practitioners and policymakers to engage each student—and especially low-income and marginalized youth—in acquiring the skills, knowledge, and expertise they need to succeed in today's and tomorrow's economy and be lifelong learners.

WWW.JFF.ORG/DEEPERLEARNING

This report was funded by The William and Flora Hewlett Foundation.

This work, *Measuring Deeper Learning: New Directions in Formative Assessment*, is licensed under a Creative Commons Attribution 3.0 United States License. Photos, logos, and publications displayed on this site are excepted from this license, except where noted.

Suggested citation: Rothman, Robert. 2018. *Measuring Deeper Learning: New Directions in Formative Assessment*. Students at the Center: Deeper Learning Research Series. Boston, MA: Jobs for the Future.

TABLE OF CONTENTS

INTRODUCTION	4
HOW FORMATIVE ASSESSMENT IMPROVES STUDENT LEARNING	4
INNOVATIVE PROGRAMS OFFER PROMISING MODELS OF FORMATIVE ASSESSMENT	7
PROMISING MODELS SHARE COMMON FEATURES	12
THE PROMISE MEETS REALITY: CHALLENGES LIE AHEAD	13
CONCLUSION: WE NEED BALANCED SYSTEMS OF ASSESSMENTS	16
ENDNOTES	17

INTRODUCTION

Assessment has become an increasingly contentious issue in education over the past two decades. Teachers, parents, and students have raised objections to the amount of testing in schools and the influence of tests on instruction. Large numbers of students have opted out of mandated tests, and districts and states have sought to reduce the number of tests they administer. Much of the objection to testing has focused on tests used for accountability purposes. But a growing chorus of educators argues not to get rid of testing but to shift the emphasis to a different type of assessment—assessments that inform instruction and learning.

For example, in 2013, the Gordon Commission on the Future of Assessment in Education, a panel of 30 leading experts in assessment and education policy led by the eminent scholar Edmund W. Gordon, recommended a much greater investment in what it called “assessments for learning: tools that provide teachers with actionable information about their students and the practice in real time.”¹ The Gordon Commission and others have called for systems of assessment that would include assessments for learning known as formative assessments in addition to assessments of learning for accountability purposes. Scholars argue the assessments should be appropriate for their intended use and should include a range of measures, from traditional on-demand tasks to complex, extended projects.² In this way, assessments, whether formative or summative, can tap a broader range of student competencies than standardized tests measure.

Over the past few years, a number of organizations have developed tools to measure these broader student competencies. They have created new models of assessments designed to inform instruction and learning, not just document the learning that has occurred. Research is showing that these measures are producing improvements in student learning.

While formative assessment is a longstanding practice in education, these models represent a departure from prior efforts in several ways. For one thing, many of them attempt to capture and measure deeper learning skills, such as the extent to which students can use knowledge to think critically and solve problems, not just memorize facts and learn procedures. In addition, many of the models use new technologies that both engage students who grew up in a

digital world and provide students and teachers with a vast array of readily accessible information about student learning.

Yet, while these efforts appear promising, they raise a number of issues that are still being debated in the field. For example:

- What is the role of students in developing and using formative assessments?
- To what extent are the tools specific to a particular subject area?
- How do the formative assessment tools fit with summative assessments used for accountability purposes?³

This paper will synthesize recent research on formative assessment, drawing from this work to elucidate its core components. It will then examine some of the new approaches to formative assessment currently being tried in schools and consider the evidence for them as well as the questions and issues they continue to raise. The paper will conclude with a look at the challenges schools and school systems face in implementing both new approaches and more established models of formative assessment.

HOW FORMATIVE ASSESSMENT IMPROVES STUDENT LEARNING

“FIRM EVIDENCE”

Formative assessment is not new. Teachers have long checked for students’ understanding and have retaught topics or presented ideas in a different way when students failed to grasp them. But the idea of systematically assessing students’ learning and providing feedback took off sharply in the late

1990s after a study by two British researchers found solid evidence of its effectiveness. In a short pamphlet and a widely read article in Phi Delta Kappa International's Kappan magazine, Paul Black and Dylan Wiliam of King's College, London, reported that they had found "firm evidence" that formative assessment practices improved student learning.

In an analysis of 43 quantitative studies of the practice, Black and Wiliam found that all of the studies "show that innovations that include strengthening the practice of formative assessment produce significant and often substantial learning gains. These studies range over age groups from 5-year-olds to university undergraduates, across several school subjects, and over several countries."⁴ Moreover, they noted, the studies show that formative assessment is particularly effective for low-performing pupils, and thus closes achievement gaps. The researchers found that the "effect size" of the gains in learning ranged from 0.4 to 0.7; a gain of 0.4, they explained, would raise the performance of an average student to the level of the top 35 percent, while a gain of 0.7 would raise the performance of a country in the middle of the distribution of 41 countries on an international assessment to the top five.⁵

Black and Wiliam caution that formative assessment is not a "silver bullet," and the results imply implementing formative assessment effectively will require significant changes in teacher practice as well as greater acceptance of the idea of student self-assessment. Nevertheless, they conclude: "There is a body of firm evidence that formative assessment is an essential component of classroom work and that its development can raise standards of achievement. We know of no other way of raising standards for which such a strong prima facie case can be made."⁶

THE IMPORTANCE OF FEEDBACK

Why does formative assessment improve student learning? John Hattie, an Australian researcher and director of the Melbourne Education Research Institute, found formative assessment works by providing feedback to students and teachers about their progress. Properly done, formative assessments alert students to what they know and can do and how this relates to their learning goals. Teachers, meanwhile, get a clear sense of where the class is in relation to these goals and what they need to do to help students advance toward them. Feedback, Hattie and his colleague Helen Timperley write, "is among the most critical influences on student learning."⁷

But not all feedback is equally effective. The most effective feedback provides information that can be used to change strategies. According to a typology of feedback developed by Hattie and Timperley, task-level feedback can tell the student and teacher how well tasks are understood and performed. However, task-level feedback is only effective if it is also related to feedback at the process level—i.e., the main processes needed to understand and perform the tasks—and/or feedback regarding self-monitoring, regulating, and directing of actions (the self-regulation level). The least effective feedback is the kind most commonly found in classrooms—personal evaluations of the learner that provide little information about how to proceed in learning.

School and classroom conditions also govern whether formative assessment will be effective. Students must have opportunities to revise their work and incorporate the feedback they receive. But that is not the case in many classrooms; students often get a grade based solely on a first draft. "You have to be able to revise based on feedback," said Heidi Andrade, an associate professor of education at the University at Albany—State University of New York. "If not, there's no use getting feedback."⁸

The type of assessment matters as well. In the early 2000s, in the wake of the No Child Left Behind (NCLB) Act, many commercial test publishers produced tests they called "formative assessments" that were designed to provide periodic checks on student performance in advance of the end-of-year tests required for accountability purposes.⁹ Lorrie Shepard argues these benchmark or interim assessments are more properly considered "formative program evaluation tools," rather than formative assessments. The data they provide is too coarse-grained to yield information on where students are struggling, and they do not provide feedback that would suggest a course of improvement.¹⁰ Shepard writes:

For most teachers, scores on benchmark tests simply signal which students are most at risk and therefore require the most attention rather than indicating the specific learning area that is in need of improvement. Such focusing of effort may indeed be one of the primary purposes for using these assessments, but the scores do not provide substantive insights about how to intervene.¹¹

In fact, Shepard argues, a teacher would have to conduct 1,000 "mini-lessons" over a course of a year to respond to everything every student missed on the interim tests.

FUNDAMENTAL COMPONENTS

What type of assessment is appropriate for formative purposes? Margaret Heritage, a senior scientist at WestEd and a leading expert on formative assessment, has identified four “core constituents” of the practice: learning goals, gathering evidence of student learning, action to close gaps, and student involvement.¹²

Identifying learning goals is the first step. Teachers and students must have a clear sense from the outset of what students are expected to learn. In some cases, the standards for student performance that all states have adopted represent these goals. However, the standards are usually written at a relatively broad level and reflect expectations for the end of each grade. Learning goals can be more specific and represent intermediate steps toward meeting the standards’ goals.

The learning goals can shape student performance. To illustrate, Andrade described an art class she observed in New York City. Students were putting together collages, but the teacher was disappointed with their work. When asked about her reaction, the teacher revealed that the learning goal stated that students should use three different types of paper, and they complied. But what the teacher wanted was for students to understand how their choice of paper could enhance what they expressed through their collages. When the teacher explained that learning goal to the students, their work improved dramatically.¹³

The second step of a formative assessment is gathering evidence about student learning. This can be done in a formal way, through an assessment task. But teachers can also gather evidence informally, by asking students to explain what they know and how they know it. Some teachers use low-technology tools like green, yellow, and red cups that students use to indicate whether they understand, have questions, or do not understand. Other teachers ask students to write down what they have learned and what they still need to learn on “exit tickets” that they complete before leaving class.

The third element of formative assessment is action. Once students and teachers have an idea of the gaps between what students understand and their learning goals, they then need to take action to close those gaps. Students can revise their work and take into account the feedback they received. Teachers can revise their instruction or reteach concepts that students failed to grasp.

Student involvement is a fourth component of formative assessment, according to Heritage. Students need to understand the learning goals and be able to monitor their own work. In this way, they develop the ability to regulate their learning—an ability needed throughout their lives.

The four constituents of formative assessment are tied together via what Heritage defines as learning progressions. Learning progressions—also known as learning trajectories or concept maps—describe the path learners take as they move from rudimentary understanding of a subject area toward increasingly complex knowledge and skills. While some of the learning progressions used in assessment systems are hypotheses about how students move along that trajectory, a number of progressions have been validated through research, particularly in science and mathematics.¹⁴

For example, the Vermont Mathematics Partnership Ongoing Assessment Project developed one learning progression showing the development of multiplicative reasoning. It illustrates how students advance from non-multiplicative strategies, such as guessing and using an incorrect operation; to additive strategies, such as repeated addition (e.g., $3+3+3+3+3=15$); to transitional multiplicative strategies (e.g., 3, 6, 9, 12, 15); to multiplicative strategies, such as doubling and halving (e.g., $16 \times 4 = 8 \times 8 = 64$).¹⁵

Using such learning progressions, teachers can determine not only whether a student got the right answer, but how the problem was solved, and what the teacher needs to do to advance the student to the next level of the progression. Students, likewise, can set the next level as a learning goal and evaluate their own performances.

A BALANCED SYSTEM

Because formative assessments attempt to gauge individual students’ progress toward learning goals and inform classroom practice, they are most useful for teachers and students. In many cases, it is difficult to aggregate the results from formative assessments to provide a picture of school or district performance. But the assessments that provide information on aggregate performance—large-scale assessment tests administered by states and districts—provide little information to inform instruction. The results of these tests usually come back well after the tests were administered, and the information provided is relatively coarse-grained compared with the information provided by formative assessments. For example, a state test might have

only a handful of questions on multiplication—too few for teachers to make judgments about student advancement on learning progressions.

For these reasons, researchers have called for balanced assessment systems that include formative and summative assessments, all based on the same set of standards. A more balanced assessment system will allow students, teachers, parents, administrators, and policymakers to get the information they need about student learning.

As Heidi Andrade, Kristen Huff, and Georgia Brooke, in their white paper, *Assessing Learning*, write:

It is necessary to contextualize student-centered assessment in a balanced system of formative, interim, and summative assessment because no one assessment process can inform students' approaches to learning, teachers' approaches to instruction, administrators' school- and district-level decisions, and policymakers' decisions about policy. For example, formative student self-assessment is highly individualized and actively engages students in regulating their own learning, but it is not particularly useful to any audience other than the student. In contrast, summative large-scale assessments provide useful information to district or state policymakers but cannot serve their intended purposes if they are individualized. Only a complete system of formative, interim, and summative assessments can be individualized, focused on learning and growth, motivating, amenable to actively engaging students in regulating their own learning, and capable of generating useful information for a variety of audiences.¹⁶

Some districts and states have attempted to create systems of assessment by using student portfolios composed of classroom work as summative measures of student learning for accountability purposes. In the 1990s, for example, Kentucky and Vermont included student portfolios in writing and mathematics as part of their statewide assessment systems. These efforts produced some improvements in instruction, but they encountered technical problems that made them less useful as accountability measures.¹⁷ After the enactment of the NCLB Act, which required states to test students in grades three through eight and once in high school, these measures were largely dropped in favor of less-expensive state tests.

INNOVATIVE PROGRAMS OFFER PROMISING MODELS OF FORMATIVE ASSESSMENT

In recent years, a number of organizations have developed new models of assessment that lend themselves to formative uses that take advantage of advances in assessment and learning science. These models provide feedback to students and teachers on their learning process and enable self-regulation, as John Hattie and Helen Timperley proposed, and incorporate the components of formative assessment outlined above by Heritage. Many also include summative components and aim to establish coherent assessment systems.

The following examples are intended to be illustrative. They suggest that the effort to place greater emphasis on assessment for learning, as called for by the Gordon Commission, is gaining some momentum.

COGNITIVELY BASED ASSESSMENT OF, FOR, AND AS LEARNING (CBAL)

CBAL is a research initiative developed by the Educational Testing Service (ETS) to create a comprehensive assessment system that includes both formative and summative components. The assessments are intended to measure what students have learned (assessment of learning), to inform instruction (assessment for learning), and to provide engaging tasks that are educational (assessment as learning).

The assessments consist of a series of tasks completed on computers that are based on a model of student competency developed from cognitive research. That is, the tasks are designed to measure student progress from initial understanding through mastery, from elementary grades through high school. In this way, students can understand what more complex work looks like and teachers can understand where students are on the trajectory toward competency.

As part of the competency model, ETS researchers have developed learning progressions in each subject area to guide the assessments.

The following is a learning progression developed for reading comprehension:

TARGET CURRICULAR AIM:	Use an Understanding of Text Structure to Enhance Comprehension of Informational Text
LEVEL 3:	Summarize text in terms of categories and details
LEVEL 2:	Infer appropriate categories from details
LEVEL 1:	Group details into appropriate categories
STARTING POINT:	Mastery of Critical Prerequisite Skills

Teachers can use these learning progressions to identify where students are along the trajectory toward the curricular aim and then adjust instruction based on the results. For example, one CBAL task asks students to conduct research on invasive plant species and to write a brochure based on the research. During the task, students have access to computers with links to web-based articles on the topic and are asked to evaluate the relevance and reliability of the articles. They then draft the brochure and receive feedback from the teacher. Finally, they revise the brochure and receive feedback on their ability to synthesize their knowledge. The computer-based assessment allows teachers to gather a great deal of data on the students' writing process. The system records each keystroke and mouse click and can tally how often students make revisions, refer to sources, or use tools such as dictionaries and thesauri.

CONNECTED STUDIOS

ConnectEd is a Berkeley, California-based organization that supports a high school-redesign model called Linked Learning that is in place in 30 school districts in California, Michigan, Texas, Ohio, Illinois, and New York. The model is designed to combine rigorous academics with technical training and real-world experience that provide college and career pathways for high school students.

ConnectEd built a comprehensive online platform (Figure 1), ConnectEd Studios, that has various features to support the development of high-quality Linked Learning pathways, including tools that provide teachers with support for developing performance assessments for students. These assessments can be used formatively, to support instruction and learning throughout the school year, or summatively, to provide information on whether students have demonstrated the competencies they are expected to master. Using the platform, teachers identify the competencies they want students to demonstrate—such as communication, collaboration, and digital literacy—and then choose a rubric (Figure 2) to assess student work and identify learning goals. The platform includes about a dozen validated rubrics developed by the Stanford Center for Assessment, Learning, and Equity (SCALE), Envision Learning Partners, and other organizations; teachers can edit the rubrics if they so choose.

The Stanford Center for Assessment, Learning, and Equity (SCALE) provides technical consulting and support to schools and districts that have committed to adopting performance-based assessment as part of a multiple-measures system for evaluating student learning and measuring school performance. SCALE's mission is to improve instruction and learning through the design and development of innovative, educative, state-of-the-art performance assessments and by building the capacity of schools to use these assessments in thoughtful ways to promote student, teacher, and organizational learning.

Once students download their work onto the platform (Figure 3), teachers can assess each student according to the chosen rubric by dragging and dropping performance indicators directly onto the work. In that way, students can see exactly where in their essays they demonstrated the desired competencies or where they fell short. (The system can also accommodate students' texts, videos, PowerPoint slideshows, Excel spreadsheets, and images.) Students then have opportunities to revise their work based on the feedback.

Business partners who support students in the career pathways also have access to the system, and can add comments to the

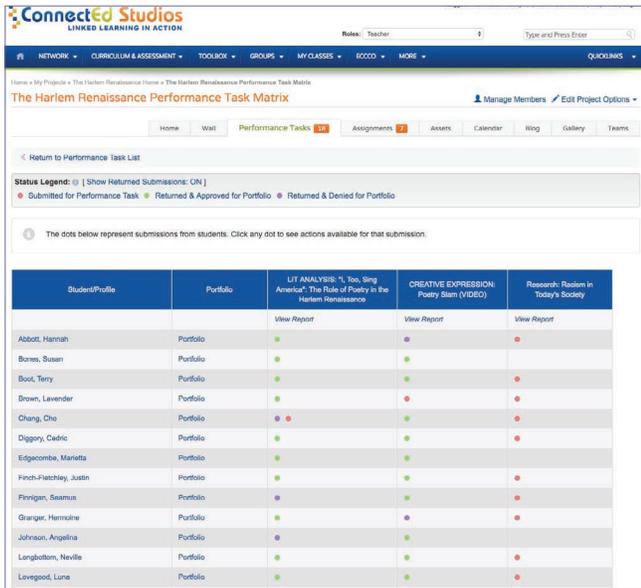


Figure 1: ConnectEd Studios Platform: The Harlem Renaissance Performance Task Matrix

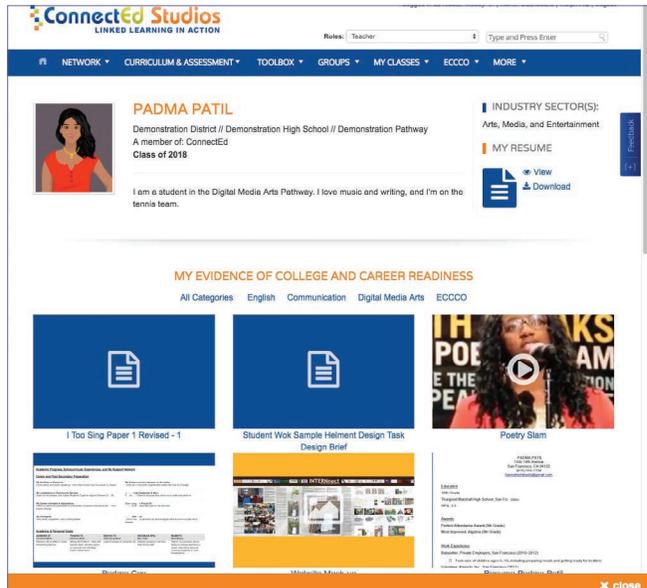


Figure 3: ConnectEd Studios Platform: My Evidence of College and Career Readiness

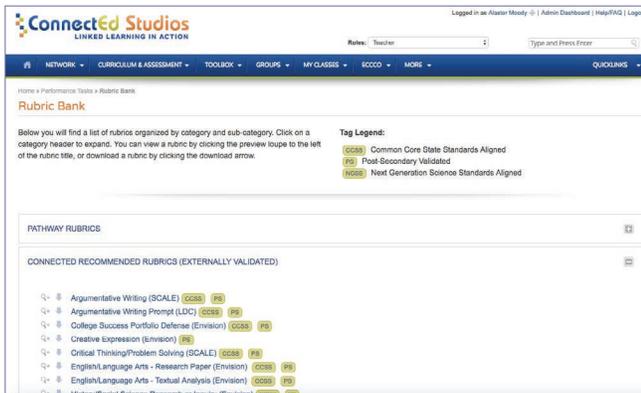


Figure 2: ConnectEd Studios Platform: Rubric Bank

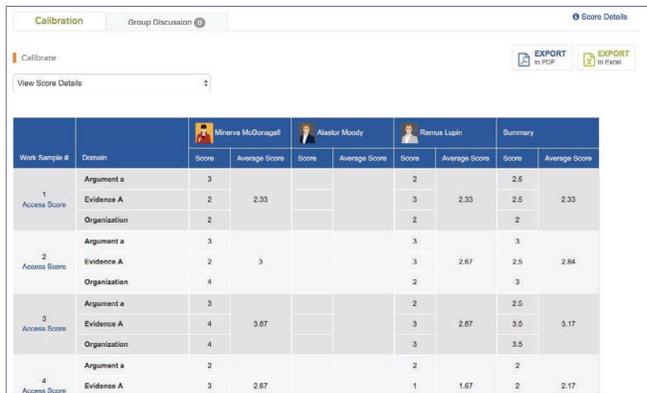


Figure 4: ConnectEd Studios Platform: Group Discussion

work and provide formative feedback to students to inform their revisions. "It's much more meaningful getting industry professionals embedded in the work at the jump," said Dave Yanofsky, director of digital learning and media for ConnectEd. "They can provide feedback on ideation and initial drafts. Once a piece of work is finalized, it's finalized."¹⁸

The platform allows teachers to see scores from all students in the class to help them understand areas they need to address and where students are struggling (Figure 4). It also allows school faculties to look at student work across classes to see where professional development for teachers is needed. A system for self-assessment and peer assessment by students is under development.

ConnectEd is also working with partner districts to support the use of assessments as summative tools. For example, some of the districts in the Linked Learning network, such as the Long Beach (CA) Unified School District, are creating digital badges that would certify whether students have demonstrated the competencies required for graduation. To support those efforts, ConnectEd Studios has developed a calibration tool that enables teachers to practice scoring student work collectively and ensure that they are using consistent standards.

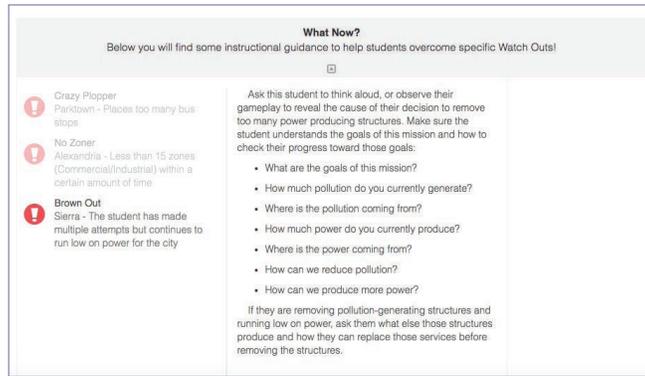


Figure 1: GlassLab Platform: What Now? Watch Out Report



Figure 2: GlassLab Platform: Shout Out Report

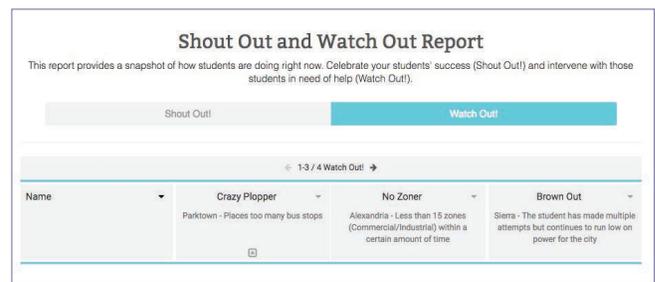


Figure 3: GlassLab Platform: Watch Out Report

GAMES FOR LEARNING AND ASSESSMENT (GLASSLAB)

GlassLab was created in 2012 as a partnership of leaders in video games, including Electronic Arts and the Entertainment Software Association, and leaders in assessment, such as ETS and Pearson, with funding from the Bill and Melinda Gates Foundation and the John D. and Catherine T. MacArthur Foundation. The goal was to develop video games that served as both learning experiences for students and assessments that would inform learning and instruction.

The first game the organization developed was a version of the popular game SimCity called SimCity EDU. The game asked students to serve as “mayor” of a city and direct its economy in ways that decouple economic growth from use of pollution-generating energy sources. A study of 400 middle school students found significant improvement in the systems-thinking abilities of students who played SimCity EDU.

GlassLab has since developed a number of additional games that teach and assess a variety of competencies,

including argumentation abilities in English language arts, understanding of ratio and proportions in mathematics, and collaborative problem-solving abilities. For each game, students and teachers receive reports indicating their competency levels as well as intervention reports (Figures 1-3) – “shout out,” “watch out,” and “what now” – that provide real-time feedback on the students’ progress. Students can then make revisions in areas flagged as “watch out” intervention reports using the “what now” information.

Paula Angela Escudra, GlassLab’s digital marketing and community manager, said the games are designed to supplement school curricula by providing students who are struggling with opportunities for engaging work, and providing enrichment to those who are doing well. “Games improve student performance by doing what games do best: dropping students into immersive environments,” she said.¹⁹ She notes that young people playing games persist even though they make mistakes, using the feedback they get to make adjustments and advance; the same process is true with games that happen to be educational.

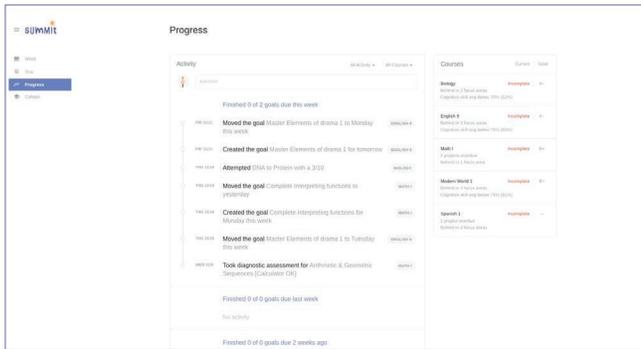


Figure 1: Summit Learning Platform: Playlist

SUMMIT PUBLIC SCHOOLS

At Summit Public Schools, an 11-school charter network in California and Washington State, each student maintains a “playlist.” (Figure 1) The playlist is an online record of work for the year. Each student begins by setting goals—such as earning certain grades or getting into certain colleges—and identifies the knowledge and skills needed to attain those goals. Students then track their progress on the performance tasks that make up the curriculum at Summit schools. Using an online platform, they can see if they met the expectations for learning that all students are expected to meet and identify where they have fallen short. They also write reflections on their progress, indicating what they need to do to improve. Teachers have access to the students’ playlists, so they can see where students are succeeding and where they need additional help.

A key element of the Summit Learning Platform, as the online tool is known, is a series of “checkpoints” (Figure 2) that take place during each project. These checkpoints represent places for students and teachers to examine evidence about their work and determine the next step. In other words, the checkpoints are used to determine whether the students can keep moving forward or whether they need to stop and regroup.

According to Adam Carter, chief academic officer at Summit, the periodic assessments are the heart of the schools’ instructional program. Unlike in traditional schools where students take assessments at the end of a unit or at the end of the year, the Summit assessments—which include written products, presentations, portfolios, and other demonstrations of knowledge and skills—are what students work on day to day. “Assessment is the main course, not dessert,” Carter said.²⁰

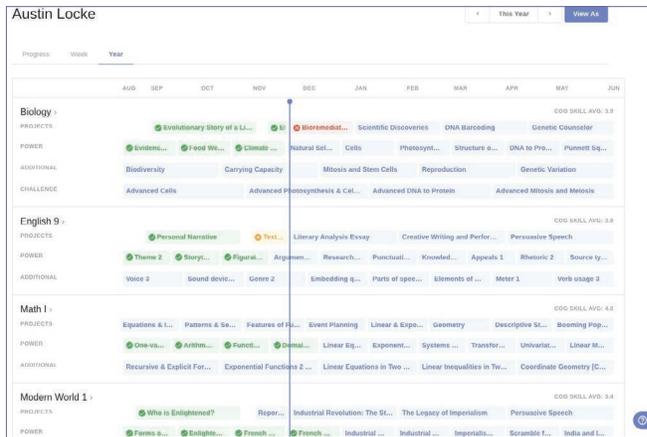


Figure 2: Summit Learning Platform: Checkpoints

To develop the assessments, Summit worked with SCALE which helped develop the measures of student progress and the rubrics for evaluating student work. The rubrics are common to all grades so all students know exactly what is expected at every point in their school career, Carter explained. “The fact that we are using the same language—textual analysis is textual analysis—is resonates with kids,” he said. “When a kid comes in, we spend a significant amount of time getting him to internalize it [the rubric]. That time pays off. And parents appreciate it—they are not getting different expectations at different grade levels.”²¹

ASSESSMENT FOR LEARNING PROJECT (ALP)

One of the most ambitious efforts to spark a new generation of assessments is a grant project funded by the Hewlett and Gates Foundations and managed by the Center for Innovation in Education (CIE) at the University of Kentucky and Next Generation Learning Challenges (NGLC). In March 2016, the initiative regranted \$3 million to 17 organizations to catalyze the development and scaling of new approaches that tap a broad definition of student success and place a stronger emphasis on assessment for learning. The grant recipients include individual schools, school districts, district consortia, and research organizations. (Summit Public Schools received a grant to expand its assessment system to include a measure of what the organization calls “habits of success,” or interpersonal and intrapersonal skills.) All of these include formative assessment tools.

Although the projects vary, most are aimed at supporting student-centered learning, providing opportunities for personalizing learning by enhancing students’ ability to

determine how their knowledge and skills are assessed, said Tony Siddall, a program officer at NGLC. “Embedded in most approaches to assessment we saw was a power dynamic that was disempowering for students,” he said. “When the manner and method of assessment is determined by adults, that puts students in the position of being receptacles of content, rather than agents. As we move to student-centered learning, we want them to have opportunities to own their own goals.”²²

For example, the Fairfax County (VA) Public Schools is piloting a project in which students design and produce a capstone project that they would then present as evidence that they have met the standards of the district’s “portrait of a graduate.” Meanwhile, Del Lago Academy in Escondido, CA, is developing a digital badging system in which students earn badges indicating competency in science and engineering. Students choose which badges to pursue.

CIE and NGLC have formed a learning community to provide a forum for the grantees to share their experiences with one another and with the broader education community. The goal is to use these experiences to inspire the field to rethink assessment, rather than to produce large-scale tools for dissemination, Siddall explains. “We try to focus more on scaling impact than on scaling individual tools,” he said.²³

PROMISING MODELS SHARE COMMON FEATURES

Although these and other new formative assessment models and projects vary in significant ways, they share some common features that suggest elements for improving instruction and learning. These include the following:

The models tap a broad range of student competencies, including deeper learning competencies.

Despite recent improvements, assessments used for accountability purposes tend to measure a relatively narrow set of competencies. The assessments seldom provide opportunities for students to conduct extended projects that ask them to solve complex, non-routine problems, or to collaborate with peers or communicate in a variety of media. Furthermore, the strong influence of accountability assessments on classroom practice has in many cases curtailed instruction that fosters attention to such learning competencies.²⁴

In contrast, the profiled formative assessment models are designed to promote instruction for deeper learning and to measure those competencies. The Summit curriculum, for example, is made up almost entirely of extended projects. Teachers start with a set of competencies that they expect all students to attain by the end of each year, and then design a series of projects that will enable students to demonstrate those competencies. Moreover, these competencies are much broader than those typically measured by end-of-year tests, and include analysis, synthesis, inquiry, and communications. The ConnectEd performance tasks develop similar competencies.

Two Rivers Public Charter School in Washington, DC, an ALP grantee, is developing assessments specifically aimed at measuring students’ critical-thinking competencies. The school is creating a set of hour-long “discipline-agnostic” performance tasks, known as exhibitions, aimed at determining how well students can transfer their critical-thinking skills from their regular classroom activities.

The game-based assessments developed by GlassLab also encourage problem solving and critical thinking. While immersing themselves in game situations, students have to identify a problem (for example, in SimCity EDU, the problem is figuring out a way to maximize energy production while minimizing pollution), make decisions about how to solve it, evaluate the solution, and then correct themselves if the solution does not work.

The student involvement in formative assessment also helps students develop the ability to self-regulate their learning, a key deeper learning competency. By providing students with feedback about their work against the standards for high quality, the assessments help students learn how to learn, said Heidi Andrade. “If we want students to learn, we’d better engage them in thinking about what counts,” she said.²⁵

The use of well-developed rubrics for evaluating student work helps ensure that the assessments measure the deeper learning competencies and can contribute to the attainment of those goals. As Randy Bennett, Distinguished Scientist in the Research and Development Division at ETS in Princeton, NJ, writes: “If the inferences about students resulting from formative assessment are wrong, the basis for adjusting instruction is weakened.”²⁶ By making clear that students are expected to develop the ability to use knowledge to think critically and solve problems, for example, the rubrics help guide instruction and learning toward those ends.

The models use technology to engage students and yield a wealth of data on student learning. Another feature the new formative assessment models have in common is their use of technology. In many cases, the assessments are completed on computers, rather than traditional pencil-and-paper tests. This provides at least two significant advantages.

First, the assessments take advantage of digital technology to make possible tasks that would be difficult, if not impossible, on paper. For example, the immersive environments in GlassLab's games enable students to manipulate environments and immediately see the consequences of their decisions. This helps them evaluate the validity of their solutions and make adjustments when necessary.

Second, computer-based assessments enable students to gain access to a wide array of materials, such as primary-source documents, and to collaborate with peers in other classrooms, states, and countries. These situations are more engaging than the often-artificial situations students face on conventional tests. Computer-based assessments also provide a vast array of information on student learning—and do so instantaneously. As noted above, in CBAL, for example, the computers can record each student's keystrokes and mouse clicks, so teachers can see what steps students took to develop their work.

While such information can be overwhelming, the platforms that organizations like Summit have created can make the assessments easier for teachers and students to use. Therefore, teachers are more likely to look at students' work and then progress together in their own professional learning, said Raymond Pecheone, the director of SCALE. "The fact that they have a platform, which is more than warehousing student work, that is dynamic and interactive, is really important," he said.²⁷

The arrays of information also help teachers identify patterns that can support their instruction and professional development. For example, the ConnectEd Studio platform helps teachers see quickly whether groups of students are struggling on a particular type of performance or whether all classrooms in the building have similar struggles. For example, if the results show that students in all classrooms tend to show little evidence of citing and refuting counterclaims when making arguments, the faculty might seek professional learning to support their ability to teach that skill.

The models enable teachers to personalize learning for each student. Teachers have long recognized that students have

unique strengths and weaknesses and learn at their own pace. But traditional school structures have made it difficult for teachers to tailor instruction to individual students. The use of formative assessments helps support personalization by enabling teachers to identify each student's progress and tailor interventions or support to specific individuals. For example, in Georgia, Henry County Schools officials are aiming to strengthen the district's ability to personalize learning by developing feedback protocols. The protocols are designed to improve the capacity of leaders, teachers, and students to analyze student work, provide effective and timely feedback, and track data collected from feedback to determine the next steps for students. The district, an ALP grantee, is also piloting a feedback process and student and teacher training using a locally developed tool called the Learner Profile in 15 pilot schools.

Formative assessments are critical to personalization because they allow students and teachers to make adjustments throughout the course of the year, rather than simply give students grades at the end of the year, said Carter of Summit Public Schools. "At the root, we are trying to make actionable, reliable, and valid measures for the purpose of intervening as rapidly as possible," he said. "We're cutting the lag time. We're not sitting back until you get an F."²⁸

This feature helps promote equity, Carter added, because it allows teachers to recognize each student's competencies and needs, rather than teach in a uniform way that leaves some students behind. "In every school I've been associated with, diversity is seen as a liability," he said. "It's superhuman to ask teachers to teach 25 kids a day who are very different. Diversity can be an asset in a learning environment, and it doesn't require superhuman effort on the part of teachers. If you know from data you have four kids struggling, you can help those kids—today."²⁹

THE PROMISE MEETS REALITY: CHALLENGES LIE AHEAD

The emergence of new methods of formative assessment is encouraging, but researchers and educators still face challenges in developing these complex tools, and teachers and school systems will face challenges of their own when it comes to implementing these programs in the classroom.

ISSUES THAT AFFECT DEVELOPMENT

While the new models appear promising, they also highlight

some of the difficulties that education professionals encounter as they try to develop effective formative assessment tools. These issues are not crippling, but they suggest that additional research is necessary to determine how formative assessments can work most effectively. The issues include:

The role of students. As noted above, student involvement is critical for formative assessments to be effective. As Margaret Heritage writes, “learning is an active, social process designed to build student independence through interaction, intervention, stimulation, and collaboration.”³⁰ To that end, assessments must provide feedback to students so that they can monitor and regulate their own learning. And as part of that process, “students must also collaborate with their teachers to determine the criteria for success for each step along the learning progression.”³¹

Developers of the new models all agree on the importance of feedback to students and the need for students to take ownership of their own learning. But many have stopped short of engaging students in determining the criteria for success. In the GlassLab games, for example, the criteria are built into the games themselves. And at Summit Public Schools and in CBAL models, the criteria—in the form of a rubric used by students and teachers to evaluate their work—were developed externally.

Carter said the Summit process engages students in their learning by enabling them to determine the criteria for success with their teachers. “There are real advantages to building a rubric with students,” he said. “But everything’s a tradeoff. Time is a valuable commodity. Is having students build a rubric a higher value than internalizing a single rubric, grade 3 through 12? Students understand the expectations and take ownership over the work of their projects.”³² He added that not all teachers are equally capable of managing the process of co-developing rubrics with students. “There are teachers—the exceptions, not the rule—who can lead students effectively through the rubric process. That’s a huge PD [professional development] lift. At scale, getting teachers to effectively manage the process is not a place we are putting our energy.”³³

Generic versus subject-specific assessments. As discussed above, most of the formative assessment practices used in schools today are home-grown and low-technology, such as colored cups and exit tickets. These practices help students reflect on their learning and provide evidence for them and their teachers about what they have learned and what

they do not understand. Students and teachers using these practices can advance student learning.

However, researchers suggest that formative assessment is more effective when it is subject-area-specific. That is, formative assessment depends on the knowledge and skills inherent in a subject area, or cognitive domain. As Bennett writes:

[T]o be maximally effective, formative assessment requires the interaction of general principles, strategies, and techniques with reasonably deep cognitive-domain understanding. That deep cognitive-domain understanding includes the processes, strategies and knowledge important for proficiency in a domain, the habits of mind that characterize the community of practice in that domain, and the features of tasks that engage those elements. ... [A] teacher who has weak cognitive-domain understanding is less likely to know what questions to ask of students, what to look for in their performance, what inferences to make from that performance about student knowledge, and what actions to take to adjust instruction.³⁴

Based on that idea, the rubrics that set criteria for student work in the CBAL and Summit models, at least, are subject specific. “You can’t be creative generally,” said Pecheone, who helped develop the Summit rubrics. “You have to be creative about something.”³⁵

The relationship between formative and summative assessments. While the new models of formative assessment were developed, at least in part, to address the perceived over-emphasis on accountability assessments, the accountability tests have not gone away. States continue to administer assessment tests to every student in grades 3 through 8 and once in high school. These assessment tests continue to carry great weight, although less so than in the NCLB era.

To maximize the effectiveness of both forms of assessments, states and districts should develop systems of assessment in which both types contribute information to different audiences at different times, based on the same learning goals. “The whole idea is that the content, format, and design of summative assessments and formative assessments should be in sync with one another, and with instruction and standards,” said Bennett. “All should be working together.” However, he added, “that’s very hard to engineer.”³⁶

The new models outlined here have tried to address this

challenge in different ways. CBAL includes both formative and summative components, all based on the same cognitive framework, but it is, at this point, a research project that is not in place on a large scale. The GlassLab games were designed to assess aspects of the Common Core State Standards and the Next Generation Science Standards, which have been adopted by numerous states. SCALE has conducted a study to show the alignment between the Summit rubric and these standards.

The California Performance Assessment Collaborative, also an ALP grantee, is aiming to help the state develop a system of assessments by influencing state policy. The consortium, a group of large districts and school networks, intends to implement performance assessments and share information about them, with the goal of identifying the supports and conditions needed to create a system in which performance assessments can serve as measures of college, career, and civic readiness. In the meantime, Summit has developed a system to provide end-of-year grades for students based on their performance in the year's projects. The school network developed the system to enable students to apply to the California State University system, said Carter. "That's not the world we want to live in—to average things out, and give students a letter grade—but we'd be putting kids at a disadvantage if we didn't." ³⁷

IMPLEMENTATION CHALLENGES IN THE CLASSROOM

Research into and development of new models and other approaches to formative assessment will continue, and the issues discussed above will be addressed and solutions can likely be worked out. But as this happens, researchers and practitioners suggest that schools face several fundamental challenges that need to be addressed to make formative assessment effective on a large scale. These challenges include:

Time. In order for teachers and students to use formative assessment to the greatest effect, teachers need to be able to pause in their instruction, gather evidence about student understanding, analyze the evidence along with students, allow students to revise their work, and adjust their instruction and reteach material if necessary. All of that takes time, and many schools have packed curricula that leave teachers with little time for these activities.

One way to address this challenge is to redesign the curriculum, as Summit did, to focus on extended projects and periodic assessments. But that approach is not feasible

in all schools. Another way is to enable teachers to rethink assessment as integral to instruction, rather than separate from it. This is a novel notion to many teachers, according to Heritage. "[T]he idea that assessment and teaching are reciprocal activities is still not firmly situated in the practice of educators," she writes. "Instead, assessment is often viewed as something in competition with teaching, rather than as an integral part of teaching and learning." ³⁸

Professional Development. Even if schools find time for teachers to implement formative assessment, many teachers lack the knowledge to develop appropriate assessments or to interpret the results, researchers say. "There is an assessment-literacy gap out there," said Pecheone. ³⁹

Some of the new models have attempted to address this challenge by making the systems so user-friendly that teachers do not need a doctorate in educational measurement to understand the results. Nevertheless, assessment results are always subject to error and need to be interpreted with care. ⁴⁰

ConnectEd is looking to support teachers by developing a set of blended learning modules to help them understand how to score performance assessments and interpret the results. The organization recognizes that not all teachers are equally capable of doing so at this point, said Yanofsky, ConnectEd's digital director. "Not everybody is a high flyer and can use [ConnectEd Studio] effectively," he said. "We want to provide scaffolding and supports." ⁴¹

But ensuring that teachers have the knowledge and skill required to implement the assessments and interpret the assessment results is not the only professional development challenge. Few teachers are able to use the results to revise their teaching to build student understanding, said Andrade. "Teachers struggle with both ends of formative assessment," she said. "They struggle with transforming standards into learning goals, and they struggle with making adaptations to their instruction." ⁴²

To help address that issue, the Center for Collaborative Education, another ALP grantee, is developing a micro-credential for teachers who demonstrate the ability to design and score high-quality performance assessments. While this effort can help, the problem is still significant, said Tony Siddall of Next Generation Learning Challenges. "Good formative assessment, and assessment for learning in general, relies much more on teachers' skills than is typically discussed," he said. ⁴³

CONCLUSION: WE NEED BALANCED SYSTEMS OF ASSESSMENTS

In its 2013 statement, the Gordon Commission argued for a greater emphasis on assessments for learning. The new models suggest that there are some promising developments on that front. However, the Commission did not call for doing away with assessments for accountability. Rather, it urged the development of “systems of assessment” in which formative and summative assessments “work together in synergistic ways.”⁴⁴

What would such a balanced system look like? As David Conley and Linda Darling-Hammond suggest in their report, *Creating Systems of Assessment for Deeper Learning*, it would consist of multiple forms of assessment that provide “information for distinctive purposes to different audiences: students, parents, teachers, administrators, and policymakers at the classroom, school, district, and state levels.”⁴⁵ In that respect, it would include large-scale tests for accountability purposes as well as classroom assessments that support instruction and learning.

The key is the word system. In a system, the assessments for different purposes are designed in a coherent fashion to complement one another. Collectively, they measure all of the competencies students need to develop to be ready for college, careers, and citizenship, and they support continuous improvement at all levels.

Top-performing nations and regions, such as Singapore and Queensland, Australia, have built coherent systems of assessment.⁴⁶ Other countries, such as Norway and Sweden, have been successful at creating systems of assessment, but those countries administer summative tests less frequently than the United States does.⁴⁷ The United States is not there yet. But the emergence of high-quality formative assessment models suggest that the nation is moving in that direction.

ENDNOTES

1. Gordon Commission on the Future of Assessment in Education, *A Statement Concerning Public Policy* (Princeton, NJ: Educational Testing Service, 2013), 5.
2. Heidi Andrade, Kristen Huff, and Georgia Brooke, "Making Assessment Student Centered," in R. E. Wolfe, A. Steinberg, and N. Hoffman, eds., *Anytime, Anywhere: Student-Centered Learning for Schools and Teachers* (Cambridge, MA: Harvard Education Press, 2013), 55-76. This chapter summarizes a research study commissioned by JFF entitled Assessing Learning.
3. The authors and editors acknowledge that this list of investigations is missing the key component of building educator capacity for new formative assessments. While this piece focuses on students, tools, and systems, the issue of capacity—and policy's role in building capacity—is investigated in far more detail in this JFF series on assessment: Pecheone et al., *Redesigning Assessment Systems: Emerging Lessons from Three States* (May 2018).
4. Paul Black and Dylan Wiliam, "Inside the Black Box: Raising Standards through Classroom Assessment," *Kappan* (October 1998), 2.
5. *Ibid.* 2.
6. *Ibid.*, 12.
7. John Hattie and Helen Timperley, "The Power of Feedback," *Review of Educational Research* 77 (2007), 81-112.
8. Personal communication, May 9, 2017.
9. Stuart Kahl, "Where in the World Are Formative Tests? Right Under Your Nose," *Education Week* 25(9) (October 26, 2005), 38.
10. Lorrie Shepard, "Formative Assessment: Caveat Emptor," paper presented at the ETS Invitational Conference, "The Future of Assessment: Shaping Teaching and Learning," New York, October 10-11, 2005.
11. *Ibid.*, p. 5.
12. Margaret Heritage, *Formative Assessment in Practice: A Process of Inquiry and Action* (Cambridge, MA: Harvard Education Press, 2013).
13. Personal communication, May 9, 2017.
14. Frederic A. Mosher, *The Role of Learning Progressions in Standards-Based Education Reform*, CPRE Policy Brief RB-52 (Philadelphia: University of Pennsylvania, Graduate School of Education, Consortium for Policy Research in Education, September 2011). http://www.cpre.org/sites/default/files/policybrief/1218_lppolicybriefwebready.pdf.
15. Philip A. Daro, Frederic A. Mosher, and Tom Corcoran, *Learning Trajectories in Mathematics: A Foundation for Standards, Curriculum, Assessment, and Instruction*, CPRE Research Report #RR-68 (Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education, January 2011).
16. See Andrade, Huff, and Brooke, "Making Assessment Student Centered," 4.
17. Ruth Chung Wei, Raymond L. Pecheone, and Katherine L. Wilczak, *Performance Assessment 2.0: Lessons from Large-Scale Policy and Practice* (Stanford, CA: Stanford University Graduate School of Education, Stanford Center for Assessment, Learning, and Equity, 2014).
18. Personal communication, May 26, 2017.
19. Personal communication, May 12, 2017.
20. Personal communication, March 17, 2017.

ENDNOTES (CONTINUED)

21. Ibid.
22. Personal communication, April 18, 2017.
23. Ibid.
24. Pedro Noguera, Linda Darling-Hammond, and Diane Friedlaender, *Equal Opportunity for Deeper Learning* (Boston: Jobs for the Future, October 2015).
25. Personal communication, May 9, 2017.
26. Randy E. Bennett, "Formative Assessment: A Critical Review," *Assessment in Education: Principles, Policy and Practice* 18(1) (February 25, 2011), 14. http://www.ets.org/research/policy_research_reports/publications/article/2011/imec.
27. Personal communication, May 19, 2017.
28. Personal communication, March 17, 2017.
29. Ibid.
30. Heritage, *Formative Assessment in Practice*, 5.
31. Margaret Heritage, "Formative Assessment: What Do Teachers Need to Know and Do?," *Kappan* 89(2) (October 2007), 140-145.
32. Personal communication, March 17, 2017.
33. Ibid.
34. Bennett, "Formative Assessment," 15.
35. Personal communication, May 19, 2017.
36. Personal communication, April 11, 2017.
37. Personal communication, March 17, 2017.
38. Heritage, "Formative Assessment," 140.
39. Personal communication, May 19, 2017. The work that organizations such as SCALE and Envision Learning Partners are doing to engage teachers in making such a fundamental shift is described in another paper in this JFF series on assessment: Pecheone et al., *Redesigning Assessment Systems*.
40. Bennett, "Formative Assessment," 15.
41. Personal communication, May 26, 2017.
42. Personal communication, May 9, 2017.
43. Personal communication, April 18, 2017.
44. Gordon Commission, *A Statement Concerning Public Policy*, 5.
45. David T. Conley and Linda Darling-Hammond, *Creating Systems of Assessment for Deeper Learning* (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2013), 6.
46. Ibid.
47. Heritage, *Formative Assessment in Practice*.

JOBS FOR THE FUTURE

88 Broad Street, 8th Floor, Boston, MA 02110
122 C Street, NW Suite 650, Washington, DC 20001
505 14th Street, Suite 340, Oakland, CA 94612

WWW.JFF.ORG

